

Kvalitetsevaluering av offentlege nettstader – kva er det vi evaluerer?

Arild Jansen og Svein Ølnes

Vestlandsforskning
Postboks 163
6851 Sogndal

Epost:, arild.jansen@vestforsk.no, svein.olnes @vestforsk.no

<http://www.vestforsk.no>

Kvalitetsevaluering av offentlege nettstader – kva er det vi evaluerer?

Samandrag:

Artikkelen rapporterer erfaringar frå eit prosjekt med evaluering av kvalitet på offentlege nettstader og drøftar ulike sider ved utforming av indikatorar og målemetodar for slike evalueringar.

Frå 2001 til i dag har meir enn 1 100 offentlege nettstader vorte evaluerte med utgangspunkt i indikatorar utvikla av prosjektet. Artikkelen drøftar vidare kvalitetsomgrepet og om det i det heile kan definerast objektive kriterium for kvalitet på slike tenester. Eit viktig spørsmål som blir diskutert, er om ei kvalitetsmessig forbetring basert på dei gitte indikatorane, nødvendigvis inneber betre brukartenester. Vidare blir det stilt spørsmål ved samanhengen mellom kvalitet og nivå på tenester i høve den såkalla tenestetrappa. Prosjektet kan visa til gode resultat ut frå målsetjingane og det kan også påvisast klare kvalitetsmessige forbetringar av offentlege nettstader målt i høve til dei definerte indikatorane.

Artikkelen konkluderer likevel med at indikatorar bør omfatta mange kriteria, og fleire enn dei som ligg til grunn for dagens målingar. Det er også nødvendig å komplettera dagens kvalitetsmålingar med andre typar undersøkingar; både for å gi eit betre bilete av total kvaliteten på ein nettstad, men også som ledd i vidareutviklinga av indikatorar.

Innleiing

Det har dei seinare åra vakse fram ei rekkje offentlege nettstader (vevtenester) som skal gjera publikum si samhandling med det offentlige enklare og meir effektiv. Dels skal desse presentera informasjon på ei enkel måte, og dels skal dei supplera og delvis erstatta menneskeleg kontakt.

Dette gjer det naudsynt å sjå på i kva grad vevtenestene oppfyller dei forventningar og krav som publikum har til desse. I samband med prosjektet *Norge.no* – etableringa av ein portal for offentlig informasjon og offentlege vevtenester – var ei tilleggs målsetjing å heva kvaliteten på offentlege vevtenester.

Statskonsult, Vestlandsforskning og internettportalen *norge.no* har saman utvikla indikatorsett for kvalitetsevaluering av offentlege nettstader. Statskonsult har vore prosjektansvarleg medan Vestlandsforskning har hatt ei rådgivarrolle og utført dei fleste evalueringane. *Norge.no* har hatt ansvaret for presentasjon av resultat og kommunikasjon med dei evaluerte verksemdene. Arbeidet starta i 2001 og sidan den tid har i alt rundt 1 200 offentlege nettstader vorte evaluerte. Det første indikatorsettet i 2001 hadde 21 indikatorar og i vidareføringa av arbeidet vart det utarbeida eit delvis nytt sett på 25 indikatorar brukt ved evaluering i 2003. Denne artikkelen vil berre ta for seg den siste evalueringa i 2003, også kalla Kvalitetsevaluering 2003, og i tillegg fokusera mest på kommunal sektor.

Av andre land er det berre Danmark som har gjennomført liknande evalueringar i stor skala. Dei danske kvalitetsevalueringane starta også opp i 2001 og har på mange måtar vore endå meir omfattande enn dei norske. Dei har i ei tilsvarande evaluering definert eit indikatorsett basert på i alt 104 indikatorar; 30 generelle (verksemdsuavhengige) og 64 spesifikke indikatorar retta mot ulike typar verksemdar (Videnskapsministeriet, 2003).

Den danske tilnærminga peiker på ei viktig problemstilling: I kva grad kan ein nytta dei same kriteria for alle typar offentlege verksemdar? Nettstader må nødvendigvis spegla dei tenestene dei tilbyr, tematisk innretning, geografi, organisatoriske tilhøve etc.

Viktige rammer for arbeidet med evaluering av kvalitet på offentlege nettstader har vore mellom anna eEurope 2005-planen (EU-kommisjonen, 2002) og den korresponderande eNorge 2005 (NHD, 2003). Frå eEurope-planen er det særleg samanlikninga (*benchmarking*) av utvalde offentlege tenester og bruken av modellen *tenestetrappa* som har vore retningsgivande.

Ei satsing på "e-forvaltning" som mellom anna desse planane er uttrykk for, inneber såleis meir samhandling mellom brukarane og vevtenestene og vidare auka spesialisering og integrasjon mellom nettstaden og andre system. I kva for grad gir det då meining å evaluere slike nettstader ved standardiserte indikatorar, slik ein til dømes gjer med tenestetrappa? Kan ein setja likskapsteikn mellom nivå i tenestetrappa og grad av kvalitet? I stor grad er det det som blir gjort i dagens målingar, både for kvalitetsevalueringa og for andre typar samanlikningar (*benchmarking*) som t.d. eEurope-statistikken.

Vi vil ikkje argumentera mot slike målingar og testar generelt, tvert om meiner vi slikt arbeid er nyttig for å betra kvaliteten på offentlege vevtenester. Men vi peikar på nokre manglar i ei slik tilnærming og vi ønskjer difor å bidra til eit betre grunnlag for dette arbeidet. Korleis skal vi forstå kvalitet og i kva for grad kan vi nytte "objektive" (kontekst-uavhengige) kriteria i måling av kvali-

tet på vevtenester? Og vidare, korleis kan ein utvida dei typar kriteria indikatorane blir forma for og ulike metodar for å supplera dagens målingar?

Denne forskinga er basert på ei metodisk induktiv tilnærming der vi tek utgangspunkt i erfaringane vi har gjort så langt. Ved å trekkje inn andre teoretiske perspektiv ønskjer vi å utvikle eit betre rammeverk for framtidige evalueringar. I neste kapittel gjer vi greie for arbeidet med å evaluere offentlege nettstader, medan vi i kapittel 3 drøftar nokre av dei utfordringane vi ser knytte til utviklinga mot e-forvaltning i høve evalueringa. Deretter presenterer vi i kapittel 4 eit breiare teoretiske perspektiv på kvalitet og kvalitetsmålingar før vi i avslutninga skisserer nokre element av eit utvida rammeverk for kvalitetsevaluering.

Data frå kvalitetsevalueringane

I dette kapitlet presenterer vi data frå kvalitetsevalueringane utførte i tida 2001 til 2003, med hovudvekt på 2003-undersøkinga og resultat frå kommunesektoren. I denne perioden er det gjennomført to undersøkingar som omfattar alle offentlege nettstader¹.

RAMMER FOR KVALITETSARBEIDET

Få ord vert meir brukte og misbrukte i dagens ordskifte enn kvalitet. Så vel offentlege som private verksemdar framhevar at dei skal auke kvaliteten på varer og tenester, men ofte utan at dei seier kva dei meiner med kvalitet. Eit par enkle definisjonar kan illustrera dette (Dahlbom og Mathiassen 1993, Braa og Øgrim 1995):

- Eit system si evne til å *tilfredsstille krav*, forventningar og ønskje som er stilte
- Forholdet mellom *forventa og opplevd* yting av eit system

Desse definisjonar legg vekt på ulike sider. Den første legg vekt på å måle skilnad mellom det som er spesifisert og det som vert målt eller registrert ut frå ”objektive” kriteria, medan den siste baserer seg på opplevde eigenskapar, dvs. *’ei subjektiv vurdering av den enkelte bruker som er berørt’*. I ISO-9000 nyttar ein denne definisjonen:

- *”Kvalitet defineres som helheten av egenskaper en enhet har og som vedrører dens evne til å tilfredsstille uttalte eller underforståtte behov (ISO-8402)”*.

Sjølv om denne definisjonen er mykje brukt og er relativt generell, er den som ISO-standardar flest mest innretta på prosessar og evaluering og kvalitetssikring av desse.

I Statskonsult, norge.no og Vestlandsforskning sitt arbeid med å betre kvalitetssikring blir ein liknande definisjonen lagt til grunn:

*Kvalitet på nettjenester betyr i dette prosjektet at offentlig informasjon og formidling av tenester på Internett må holde en predefinert standard eller et nivå som kan tilfredsstille en del sentrale brukerbehov.*²

¹ Alle offentlege nettstader er i prosjektet definert som alle kommunale og fylkeskommunale nettstader samt rundt 200 statlege verksemdar valde ut i samråd med Statskonsult og norge.no med utgangspunkt i andre IT-relaterte undersøkingar.

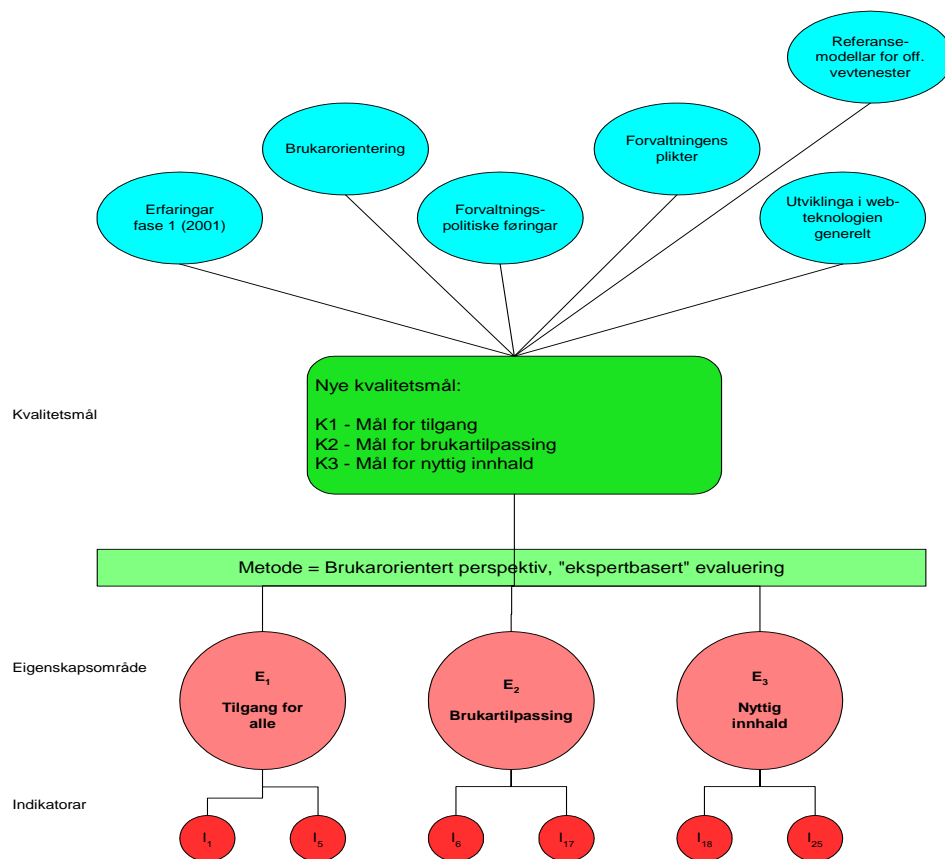
² <http://www.kvalitetpaanett.net/OmKvalitet.htm>

Dei sentrale rammefaktorane for utvikling av kvalitetsindikatorar for offentlege nettstader var:

- Politikkområde (sentral IT-politikk (AAD 2002), statleg informasjonspolitikk, eNorge 2005 og eEurope 2005)
- Forvaltningsinterne prinsipp (Forvaltningslova, Offentleglova, kompetanse og ressursar i forvaltninga)
- Referansemodellar/utviklingsmodellar (tenestetrappa, sjå Figur 3)
- Allment aksepterte standardar og retningslinjer på verdsveven (W3C-tilrådingar generelt, WAI - *Web Accessibility Initiative* - spesielt)

I tillegg vart det lagt stor vekt på intern læring i arbeidet med kvalitetsindikatorar, basert på erfaringar og tilbakemeldingar frå første runde i evalueringsarbeidet høyringsfråsegner og diskusjonar basert på forslag til nytt indikatorsett. Vidare var det krav om at resultatane skulle kunna etterprøvast og ei ekstern evaluering basert universelle indikatorar som skulle passa alle forvaltningsnivå og – sektorar. Det vart søkt å utvikle normative kvalitetskrav som kunne fungera som retningslinjer for praktisk arbeid med vevtenester. Eit godt eksempel på at dette har fungert, er at mange har brukt indikatorsettet som del av kravspesifikasjonen overfor leverandørar. Endeleg har eit brukarorientert perspektiv lege til grunn for evalueringane. Rammene for arbeidet kan skisserast som i figuren under:

Vurdering av kvalitet på offentlege vevtenester



Figur 1: Rammer for arbeidet med utforming av kvalitetsindikatorar (Ølnes, 2003, side 11)

METODE FOR EVALUERING AV NETTSTADER

Evaluering er eit omfattande fagområde og det eksisterer mange metodar knytt til ulike former for evaluering. Jennifer Preece skil mellom fire hovudtypar evalueringar i samband med design av brukargrensesnitt (Preece 2002):

1. *"Quick and dirty"* – rask og relativt usystematisk tilbakemelding frå brukarar, kollegaer og andre
2. *Usability testing* – brukartestar der brukarar blir plasserte i kontrollerte omgivnader og observerte når dei brukar ei teneste som skal evaluerast
3. *Field studies* – feltstudiar der brukarar blir observerte i sine naturlege omgivnader og korleis dei då reagerer
4. *Predictive evaluation* – forutsigbar evaluering; ein evalueringmetode som ofte gjer bruk av heuristiske metodar³ for å finna brukarproblem (*usability problems*). Det sentrale med metoden er at brukaren ikkje treng vera til stades under evalueringa. Vi kallar slike evalueringar ofte ekspertvurderingar/-evalueringar.

Evaluering av kvalitet på nett-tenester kan i prinsippet utførast på fleire ulike måtar, t.d. ved

- eigenevaluering (dei som har ansvar for vevtenestene)
- ekspertevaluering (ei gruppe godt trene evaluatarar)
- brukartestar og brukarundersøkingar (brukarane sjølve)

Av desse vart altså ekspertevaluering valt som metode og den samsvarer med det Preece kallar *Predictive evaluation*. Grunnen til det var først og fremst eit ønskje om konsistent evaluering og ei evaluering som var ressursmessig forsvarleg. Ei eigenevaluering ville kunna by på konsistensproblem ved at det ikkje vart gjort like evalueringar av dei same indikatorane. Det kunne blitt systematiske skilnader på grunn av at ulike personar sto bak og desse personane også til dels kunne ha eigeninteresse av å påverka resultatet. Ei diplomoppgåve ved Norges Informasjonsteknologiske høgskole, (NITH 2002), viser at dette lett kan skje. I oppgåva vart indikatorsettet for kvalitetsevalueringa drøfta ut frå heuristiske prinsipp formulerte av Jakob Nielsen (Nielsen 1993). I oppgåva vart det konkludert med at skilnader i evalueringresultat kunne forklarast ut frå ulike evaluatarar.

Usability

Jacob Nielsen drøftar i boka *Usability Engineering* (Nielsen 1993) 'brukssidene' ved eit system, og refererer til omgrep som 'brukarvennlegheit' (user friendliness), 'bruksnytte' (usability), 'grad av nytte' (usefulness), osv, som alle kan sjåast på som ulike dimensjonar ved omgrepet 'akseptbarheit' (system acceptability). Han vel derfor å nytte 'usability' og knyter det til desse eigenskapar: i) lett å lære ii) lett å bruke iii) lett å hugse iv) få feil, v) opplevd tilfredsstillande (subjectively pleasing). Eit av hovudpoenga hans er at ulike brukarkategoriar og brukssituasjonar og vidare individuelt ulike preferansar gjer brukartesting (*usability testing*) vanskeleg. Han peiker på 3 hovuddimensjonar:

- Røynsle med datamaskiner og relevante datasystem generelt
- Røynsle med dei konkrete systema (nybegynnar – ekspert)
- Kunnskap og kompetanse på det konkrete fagområdet systemet skal brukast på

³ Med heuristisk metode forstår vi å løysa eit problem ved å evaluera erfaringar og ved å prøva og feila til ein har ei løysing.

Den heuristiske evalueringa gjeld særleg innafor det generelle området brukarnytte/-vennlegheit/-tilpassing (på engelsk *Usability*), og Jakob Nielsen har formulert 10 heuristiske prinsipp for *usability* (Nielsen 1993).

Nielsen peiker derfor på problema med brukertesting generelt der resultatane vil sprike på bakgrunn av korleis (test)brukarane plasserer seg på dei 3 dimensjonane overfor. Vi er her samde med Nielsen og meiner det er naudsynt at brukartesting tek omsyn til alle tre dimensjonane.

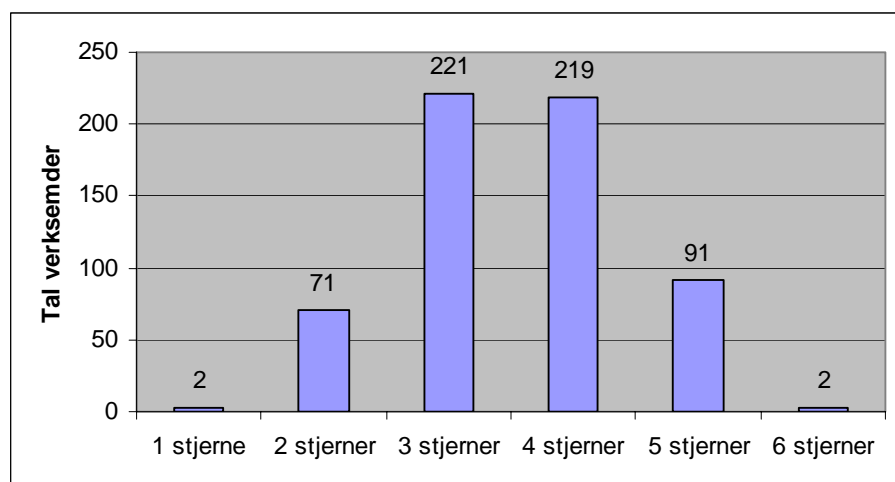
Slik brukartesting kunne vore ein alternativ metode for å sikra eit reelt brukarperspektiv på evalueringa, men måtte forkastast ut frå ressursane som stod til disposisjon. Det ville likevel vore interessant å sjå på resultatet av ei brukartesting opp mot ekspertvurderinga av same tenesta. Ei slik vurdering vart gjort i samband med dei danske evalueringane og synta at det ikkje alltid var samsvar mellom ekspertevalueringa og resultat frå brukartesting av den same tenesta⁴.

GJENNOMFØRING AV EVALUERINGA

I Kvalitetsevalueringa 2003 vart det utarbeida i alt 25 indikatorar; ein del var uendra frå 2001, nokre var endra litt og ein del var heilt nye. I alt vart i overkant 600 offentlege nettstader evaluerte (200 statlege, resten kommunale).

Ei gruppe på 4-5 evaluatarar gjekk gjennom alle nettstadene og gav poeng frå 1 til 4 for kvar indikator. Enkelte indikatorar var av typen ja/nei; her vart det gjeve 1 eller 4 poeng. Andre indikatorar brukte heile skalaen. Gjennomsnittleg poengsum for dei 25 indikatorane vart til slutt rekna om etter ein 6-trinns skala og presentert som "tal stjerner".

I tillegg til evaluatarane si evaluering var det lagt opp til ei kvalitetsikring med diskusjon og re-vurdering av tvilstilfelle. På førehand var det gjennomført trening og samanliknande evaluering for at evaluatarpanelet skulle vera mest mogleg samstemt. Resultatane viste tilnærma normalfordeling både i 2001 og 2003, begge gonger med midtpunktet for kurven nokså nær midtpunktet av skalaen.



Figur 2: Totalfordeling etter kvalitetsevaluering av 606 offentlege nettstader i 2003 (Ølnes, 2003, side 21)

⁴ I Danmark vart det gjort ei brukartesting av den kommunale vevtenesta som kom best ut i ekspertvurderinga. Brukertesten viste at tenesta gjennomgåande var lite brukarvennleg.

For begge undersøkingane galdt at statlege nettstader i gjennomsnitt oppnådde høgare karakter enn kommunale⁵.

Ei direkte samanlikning mellom resultatet frå dei to undersøkingane er ikkje muleg sidan indikatorsetta ikkje er like. Men for dei indikatorane som var uendra frå 2001 til 2003, såg vi ei forbetring i resultat på i overkant av 10%. Kommunale nettstader synte relativt sett størst forbetring (+ 14%) i høve statlege nettstader (+ 7%) som tabellen under viser.

Tabell 1

Endring i resultat frå 2001 til 2003 for like indikatorar

Sektor	Endring 2001 – 2003 (%)
Statlege verksemder	+ 6,9
Kommunar	+ 14,3
Fylkeskommunar	+ 11,6
Totalt	+ 10,6

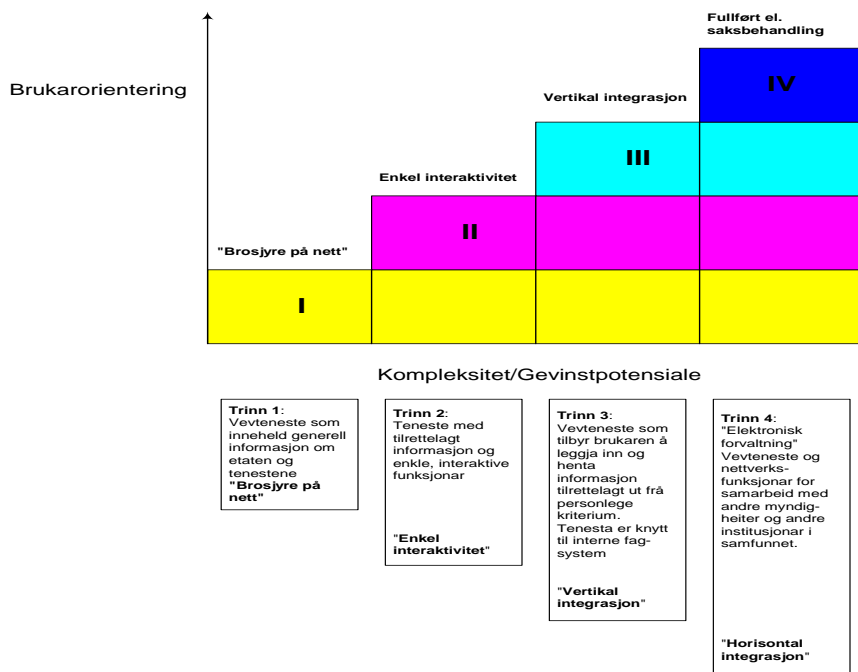
Note: Tabellen er henta frå Ølnes, 2003, side 22

Etter evalueringane i 2001 vart det gjort mange endringar i dei evaluerte tenestene, særleg hjå kommunane, og med etterfølgjande ønskje om ny evaluering. På bakgrunn av tilbakemeldingane kan vi seia at ein stor del av dei offentlege vevtenestene gjorde forbetringar sett i høve til dei formulerte kvalitetsindikatorane. Dette inntrykket vart også stadfesta av ei evaluering av den offentlege portalen *norge.no* der over halvparten av dei spurde vevansvarlege (*webmaster*) sa dei hadde teke omsyn til kvalitetsindikatorane 2001 og gjennomført forbetringar på eiga teneste.

⁵ Signifikant (95% konfidensintervall) skilnad i kvalitet mellom statlege og kommunale nettstader

Frå informasjonsformidling til spesialiserte tenester

Som påpeikt ovanfor har vi ei utvikling mot meir spesialiserte tenester som er integrerte med ulike fagsystem i dei enkelte institusjonane og som tilbyr utstrekkt samhandling mellom brukar og vevteneste. Den såkalla referansemodellen for offentlege vevtenester (Statskontoret, 2000), eller *tenestetrappa*, er brukt i ei rekkje samanhengar og har vore presentert i mange publikasjonar.



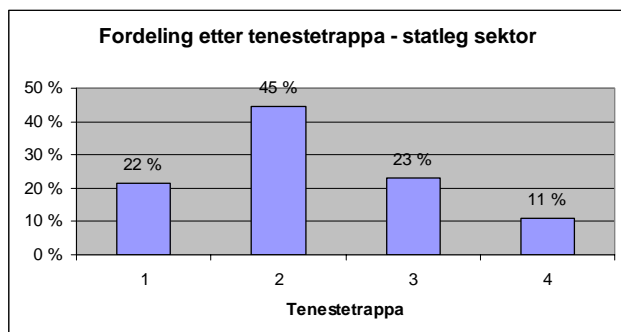
Figur 3: Tenestetrappa – ein utviklingsmodell for offentlege vevtenester (Ølnes, 2003, side 12)

Tenestetrappa syner at medan tenester på dei lågaste trinna rettar seg mot uthenting av informasjon eller enkle interaktive tenester, så vil tenestene på dei øvste nivåa i aukande grad verte integrerte i utføringa av oppgåvene.

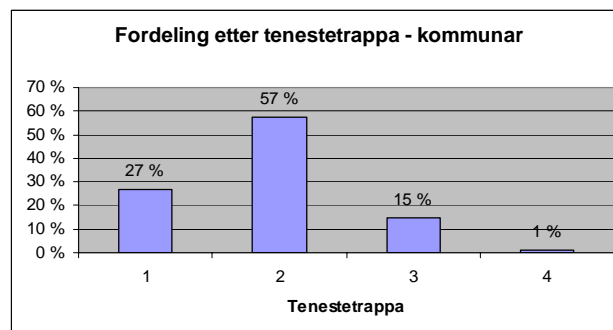
Modellen *Tenestetrappa* har gått gjennom ei endring frå den først vart lansert og slik den i dag vert brukt mellom anna i eEurope-målingane. I tidlege referansar til modellen (ANOA 1999 og Statskontoret 2000) var trinn 3 og 4 definerte som *vertikal* og *horisontal* integrering. Vertikal integrering vart definert som integrasjon mellom vevteneste og fagsystem *internt* i organisasjonen, medan horisontal integrering vart definert som integrasjon mellom vevteneste/fagsystem i organisasjon og *eksterne* tenester/organisasjonar. Det eksterne perspektivet var viktig for trinn 4. Seinare har dette vorte modifisert slik at ein i eEurope (og eNorge) i dag snakkar meir om datafangst inn i interne fagsystem og personleg tilrettelegging på trinn 3, og fullført elektronisk saksbehandling på trinn 4. I det siste tilfellet spelar det ikkje nokon rolle om saksbehandling berre er intern eller om den også involverer eksterne tenester/organisasjonar. Likevel finn vi framleis tidlegare tolkingar av tenestetrappa, mellom anna i Arbeids- og administrasjonsdepartementet sin nyleg framlagde *Strategi for IKT i offentlig sektor* (AAD 2003).

Tenestetrappa peiker på problemet knytt til forståing av korleis arbeidsoppgåver blir utførte. Det er eit uttalt mål at offentlege tenester skal 'gå' oppover trappa, det vil seia at dei blir meir komplekse

både gjennom tettare integrering med interne fagsystem og gjennom å involvere fleire etatar. I 2003 var fordeling etter tenestetrappa som følgjer i statleg og kommunal sektor:



Figur 4: Fordeling av statlege nettstader i høve tenestetrappa



Figur 5: Tenester på kommunale nettstader i høve tenestetrappa, kommunar

I eNorge-planen er modellen også sentral for måling av utviklinga i utvalde offentlege tenester på nett. Det er på grunn av rapportering til eEurope-programmet, der tenestetrappa er ein viktig modell for benchmarking av offentlege tenester i EU-landa. Også i andre undersøkingar, jfr. til dømes Accenture sine internasjonale undersøkingar om e-forvaltning⁶ ligg variantar av tenestetrappa til grunn for målingane. Tenestetrappa, og variantar av denne, er såleis den viktigaste samanlikningsmodellen for offentlege tenester på nettet internasjonalt sett.

Drøfting av kvalitet og vevtenester

Målsetjinga for prosjektet kvalitetsevaluering av offentlege nettstader var å bidra til å auka kvaliteten på offentlege nettstader. Ut frå observert endring frå 2001 til 2003, for dei indikatorane som let seg samanlikna, kan vi seia at kvaliteten har auka sett med utgangspunkt i tilpassing til indikatorsettet. Også dei mange forbetringane og re-evalueringane 2003 er eit uttrykk for betre tilpassing til indikatorsetta og dermed ei kvalitetsforbetring sett i høve til indikatorane.

Prosjektet kan på denne måten seia at målet med betre kvalitet på offentlege nett-tenester er nådd. Men er det sikkert at dei offentlege tenestene har vorte betre for brukarane? Og i kva for grad gjeld dette tenester på dei høgste nivåa i tenestetrappa?

Som døme vil vi sjå på informasjons- og tenesteportalen LivsIT (*Livssituasjonsbasert IT-system*) og er eit prosjekt drive av Statskonsult. Målet med prosjektet er å utvikla eit informasjonssystem som integrerer informasjon frå ulike forvaltningsnivå og –sektorar og presenterer denne på ein stad – kommunen si heimeside. For å få til dette, gjer LivsIT bruk av metadata og gjennom eit tematisk, livssituasjonsbasert emnehierarki blir informasjon på ulike nivå klassifisert. Systemet er omtala i ei evaluering utført våren 2003 (Ølnes og Aaberge, 2003).

Ei teneste som LivsIT er litt vanskeleg å plassera i tenestetrappa; den har element av både trinn 2 (rein informasjonsteneste) og trinn 4 (integrasjon på tvers). Når LivsIT eventuelt også blir bygt ut til å handtera tenester, vil den etter kvart bli ei teneste på trinn 4. Tradisjonelle kvalitetsindikatorar vil ha problem med å fanga opp aspekta ved ei slik teneste. Det er fordi mange aktørar er involverte

⁶ Jf t.d. Dagsavisen 29. 04. 2003: Jumboplass i IT-bruk. Norge har ramlet helt ned på 16. plass i rangeringen over hvilke land som er best på å betjene borgerne elektronisk. Danmark er nå best i Europa.

i ei teneste som til slutt blir presentert på *ein* stad for brukarane. Kvaliteten på tenesta vil vera avhengig av innsatsen i fleire ledd. I eit tenkt tilfelle kan ein brukar gå inn på bustadkommunen si nettside og via LivsIT finna fram til informasjon og tenester innan livssituasjonen til dømes ”barnefødsel”. Her vil det vera informasjon både frå kommunen og frå andre offentlege verksemder; både lokale, regionale og statlege. Om eit departement har gjort ein dårleg jobb med tilrettelegging av informasjon for LivsIT, vil det bli avdekka som ein mangel ved denne kommunen si vevteneste, medan problemet ligg ein heilt annan stad. Dette set store krav til utforming av indikatorar og dagens modell er neppe tilstrekkeleg for ei fullgod løysing.

Vi meiner at tenestetrappa er for enkel, og trur det er nausynt å sjå kvalitet ut frå fleire dimensjonar, som til dømes *omfang (kompleksitet)* som gjeld både tal ulike tenester nettstaden dekkjer og informasjonsomfang innan desse tenestene, *grad av vertikal integrasjon* med aktuelle fagsystem (til dømes grad av automatisert saksbehandling innan ei offentleg verksemd, dvs trinn 3 i tenestetrappa ovanfor), vidare *grad av/(omfang av) samhandling mellom verksemder*, tilsvarande horisontal integrasjon i tenestetrappa. Eksempel på det siste finn ein mellom anna i Brønnøysundregistra⁷, samordna opptak og i nokon grad innan Skatteetaten⁸.

Vi meiner at ein nettstad må evaluerast langs desse dimensjonane kvar for seg, og ikkje som i tenestetrappa der ein berre måler kva for trinn tenesta kan seiast å vere på og set likskap mellom trinn og kvalitetsnivå. Kvalitet må oppfattast som uavhengig av nivå på tenestetrappa og såleis at kvart nivå på tenestetrappa kan ha god eller dårleg kvalitet.

Andre sider ved det offentlege sitt ansvar bør også trekkast inn. Schartum (2002) peiker til dømes på at offentlege tenester i stor grad er knytte til å yte lovfesta tenester eller å sikre fastlagde rettar, til dømes personvern, rettstryggleik, innsyn mm. Han hevdar at ei vevteneste som er knytt til offentleg sakshandsaming må gjere greie for dei rettar dette omfattar og korleis dette er handtert, og at dette difor må inngå i evalueringa som ein viktig del av ei kvalitetsvurdering.

KVALITET ER EIT MANGESIDIG OMGREP

Kvalitet vert ofte sett på som ein objektiv storleik, slik som til dømes i software engineering tradisjonen (Sommerwille 2000), eller i ei såkalla ’hard’ system-tenking. ISO 9000-3 standardane legg og ei slik forståing til grunn (Braa og Øgrim 1995). I motsetnad til dette synet kan ei sjå kvalitet frå eit subjektivt perspektiv, grad av tilfredse hos brukarane (Kim 1989) eller samsvar med føremålet (Kitchenham 1989).

Dahlbom og Mathiassen(1993) peiker på at eit grunnleggjande problem i arbeidet med kvalitet er ulike grupper sine oppfatningar av kvalitet, t.d. vil *brukar og utviklar* kunna ha ulike oppfatningar om kvalitetskrava, det same gjeld mellom *brukar og tilbydar*, t.d. når det gjeld prioriteringa av funksjonalitet versus brukarvennlegheit, og at omgivnaden påverkar oppfatningar av kvalitet. Også innan ein einskild organisasjon vil ulike individ eller grupper leggje vekt på ulike sider, til dømes brukarvennlegheit versus tryggleik.

Braa og Øgrim (1995) hevdar at kvaliteten på eit informasjonssystem må evaluerast ut frå minst 5 ulike aspektar: *teknisk kvalitet, brukarvennlegheit, estetiske og symbolsk kvalitetar og organisatorisk kvalitet*. Dette er i samsvar med Dahlbom og Mathiassen (op.cit) på at medan mange kvalitetskrav er knytte til eit ’rasjonalitets-perspektiv’ (menneska som målorienterte, og nytteorienterte) med vekt på funksjonalitet og effektivitet, så er meir subjektive eigenskapar viktige for oss, men

⁷ www.brreg.no

⁸ www.skatteetaten.no

dei er vanskelegare å måle: 'Quality cannot be measured, it can only be experienced'. Medan estetiske eigenskapar er knytte til visuelt inntrykk, arkitektonisk stil osv, vil symbolske eigenskapar vera berarar av kulturelle verdiar i ein organisasjon, til dømes det å vera moderne eller å vera tradisjonell, til dømes gjennom dei visuelle uttrykka ein brukar i den grafisk utforminga av nettstaden. Likeins kan ein tala om det politiske innhaldet i systema, til dømes som berarar av verdiar knytte til makt, likestilling, likeverd etc. Ein nettstad kan til dømes indirekte formidla bestemte syn på kjønnskilnader eller ikkje vera nøytral i spørsmål knytt til visse innvandrargrupper, etc.

MÅLING ELLER EVALUERING AV KVALITET

Vi har ovanfor peikt på ei rekke ulike dimensjonar og kriteria som bør leggjast til grunn i arbeidet med å betre kvaliteten ved offentlege nettstader. Dette krev at ein nyttar meir enn ei tilnærming i arbeidet, og vi vil her nemne nokre:

1. Måling av kvalitet

Dette vil seie å undersøkje om tenestene er i samsvar med kravspesifikasjonane, til dømes når det gjeld eigenskapar som tekniske krav, arkitektur, og likeins verifisere funksjonalitet, korrekte, konsistens mm. Dette blir oftast gjort med tradisjonelle software engineering metodar, kor ein i arbeidet nyttar bestemte (objektive målekriteria (metrikkar), sjå til dømes (Sommerville 2001), Software Engineering Institute (SEI)⁹.

Vi meiner ei slik tilnærming er naudsynt for kvalitetsikre dei einskilde tenestene, men og for å analysere eigenskapane ved heile nettstaden; korleis han er bygd opp, som ein gjer ved å studere strukturen, samanheng mellom de ulike komponentane, kontroll av konsistens, og forhold til lover og regler. Det er i stor grad denne metoden som er nytta i kvalitetsevalueringa referert tidlegare i artikklen

2. Brukartestar

Dette vil seie å evaluere korleis systemet fungerer i bruk for ulike brukargrupper, som inneber å involvere brukarane i evalueringa gjennom testar og spørjeundersøkingar. Slike evalueringar må vere basert på brukarane sine meir subjektive oppleving av systemet, gjennom at ein søker å avdekkje alle sider ved ein artifakt som ein trur kan vera viktig for brukaren, som også omfattar brukarvennlegheit, estetiske og symbolske eigenskapar. Her vil ein måtte ta omsyn til brukarane si erfaring med datamaskinar, røynsle med dei einskilde systema og kompetanse på det konkrete fagområdet. Men her er det og ei rekke problem

- Kva er det brukarane faktisk treng?
- Ulike partar har ulike oppfatningar om kvalitetskrav og deira prioritet, og enkelte kvalitetskrav kan stå i motstrid til kvarandre, som til dømes brukarvennlegheit versus tryggleik
- Ikkje alle brukarkrav let seg enkelt måle
- Omgivnaden påverkar oppfatningar av kvalitet

3. Forankring i organisasjonen

Her må ein undersøkje om tenestene på nettstaden samsvarar med systemet sin plass og rolle i organisasjonen, det vil seie korleis dei ulike funksjonar og tenester er forankra i organisasjonen sine rutinar og ansvarsforhold. Dette reiser og vanskelege metodiske spørsmål om korleis ein faktisk kan evaluere dette på gode måtar utan å 'gå inn i' organisasjonen. Dei danske kvalitets-evalueringane har i tillegg til målingar av den første typen også inkludert ei spørjeundersøking til

⁹ SEI: Software Engineering Institute at Carnegie Mellon University, www.sei.cmi.edu

vevansvarlege ved dei same nett-tenestene. Då får dei fanga opp delar av organisatorisk forankring, ressursbruk, driftsopplegg osv.

Vi hevdar at for å evaluere komplekse tenester som t.d. LivsIT på ein god måte, må ein utforme testar og indikatorar som er tilpassa ulike type nettstader. Det betyr at indikatorane ikkje kan vera like for alle typar vevtenester, men bør setjast saman av eit felles sett for alle pluss typespesifikke indikatorar til grupper av tenester.

Evaluering av offentleg vevtenester – får vi dei svara vi ber om?

Vi har i det føregåande drøfta ulike tilnærmingar til å måle eller evaluere eit system på, og hevdar at ein einskild metode eller tilnærming ikkje er dekkjande. Vi peikar på at det er mange ulike dimensjonar, til dømes kompleksitet, integrasjon med verksemdene sine funksjonar, samvirke med andre verksemdar, som vanskeleg blir fanga opp i dagens indikatorsystem.

Evalueringa må og omfatte mange typar kriteria, som blant anna teknisk kvalitet, arkitektur og struktur, funksjonalitet, brukarvennlegleik, estetiske og symbolske kvalitetar, organisatorisk forankring og samsvar med aktuelle lover og regler. Ein bør ta omsyn til dei fleste av desse i eit evalueringssopplegg, men ikkje leggja vekt på dei likt for ulike typer nettstader. Til dømes vil dei ulike trinna i tenestetrappa krevja ulike opplegg, og ei vevteneste med svært avgrensa funksjonsområde vil vera mykje enklare å evaluere samanlikna med ein portal med eit vidt funksjonsområde. Ein må nytte ulike tilnærmingar, både meir formaliserte metodar som målar samsvar med kravspesifikasjonar, og å teste eller evaluere funksjonen til eit system i høve til omverda, som til dømes i dei tradisjonelle brukartestar eller spørjeundersøkingar

Slike evalueringar må baserast på brukarane sine meir subjektive oppleving av systemet,

Som oftast må ein nytte begge tilnærmingane i ei evaluering, men tilpassa den einskilde tenesta. Vi meiner den einaste måten å finna ut om tenestene faktisk blir betre for brukarane ved å følgja dei oppsetta kvalitetsindikatorane, er ved å gjera ulike formar for brukarundersøkingar. Slike undersøkingar kan vera reine spørjeundersøkingar, intervju eller brukartesting. Vi meiner vidare at ei utvikling av indikatorsett bør skje i nært samspel med brukarundersøkingar, då helst gjennom

- Ulike formar for brukartesting
- Utprøving gjennom t.d. eksperimentering og prototypar
- Utarbeiding av meir detaljerte brukscenarier for å avdekkje om tenestene samsvarer med systemet sin plass/rolle i organisasjonen.

Desse ulike perspektiva vil ha konsekvensar for både korleis vi fastset kvalitetskrava, og korleis dei kan målast. Eit opplegg som tek omsyn til alle desse ulike dimensjonane vil gjere evalueringane svært tidkrevjande og lite eigna til samanlikningar på tvers av tenestetypar og funksjonsområdar, slik som slike tradisjonelle benchmarking siktar mot. Men det er viktig å vere klar over dei svake sidene dei standardiserte opplegga har. Vi treng røynsle med evaluering av komplekse offentlege nettstader og trur det ville vere nyttig å forske meir systematisk på kor gode og effektive dei ulike formene for brukartenester er.

Arild Jansen er Dr. Scient. i informatikk ved Universitetet i Oslo fra 1999. vhandlingens tittel er *Utkanten i den globale landsby. Integrasjon eller identitet?* Han er førsteamanuensis ved Avdeling for forvaltningsinformatikk, Universitetet i Oslo og vitenskapelig rådgjevar ved Vestlandsforskning. Aktuelle forskningsinteresser er IKT og regional utvikling, i nformasjonsinfrastrukturer og utvikling av offentlig forvaltning ved IKT.

Svein Ølnes er cand. agric. (NLH 1984) og forskar ved Vestlandsforskning sidan 1996. Han arbeider i hovudsak med problemstillingar knytte til IKT og offentlig forvaltning. Av nyare, større arbeid er teknisk ansvar for etablering av den offentlege informasjonsportalen norge.no, kvalitetsevaluering av vevtenester og elles arbeid i tilknytning til integrasjon av informasjon og bruk av metadata (mellom anna gjennom evaluering av metadata-prosjektet *LivsIT* og bruk av søkemotor for effektiv kategorisering av offentlig informasjon).

REFERANSAR

Arbeids- og administrasjonsdepartementet (AAD) (2002): *Fra ord til handling*, Regjeringa si IT-politiske utgreiing, Oslo.

Arbeids- og administrasjonsdepartementet (AAD) (2003): *Strategi for IKT i offentlig sektor – Strategi 2003 - 2005*, Oslo.

Australian National Audit Office (ANAO) (1999): *Electronic Service Delivery*, ANAO-rapport 18/1999-2000, www.anao.gov.au

Braa, K. og L. Øgrim (1995): “Critical View of the Application of the ISO Standard for Quality Assurance”, Avison & Fitzgerald (1995) *Information Systems Journal*

Dahlbom, B. and L. Mathiassen (1993): *Computers in Context*, Basil Blackwell

EU-kommisjonen (2002): *eEurope 2005 Action Plan*, EU Commission, http://europa.eu.int/information_society/eeurope/news_library/eeurope2005/index_en.htm

Holmane, Vidare (1999): *Kommuner og lokalsamfunn på Internett*, Kommuneforlaget, Oslo

Kim, K. K. (1989): “User Satisfaction: A Synthesis of Three Different Perspectives”, *Journal of Information Systems*.

Kitchenham (1989): “Software Metrics”, in Rook, P. (ed) *Software Reliability Handbook*, Elsevier

Nielsen, Jakob (1993): *Usability Engineering*, Morgan Kaufmann, San Francisco

NITH (2002): *Kvalitetsevalueringen av offentlige tjenester på Internett – en indikasjon på brukbarhet?*, Diplomoppgåve ved Norges Informasjonsteknologiske Høgskole, NITH, Bærum

Nærings- og handelsdepartementet (2003): *eNorge 2005*, NHD, <http://www.enorge.org>

Preece, Jennifer (2002): *Interaction Design – beyond human-computer interaction*, Wiley, New York

Schartum, Dag Wiese (2003): *Krav til åpenhet på nettsider i offentlig forvaltning. Forslag til åpenhetskriterier basert på evaluering av Larvik kommunes nettsider*. Rapport. Avd. for forvaltningsinformatikk, Universitetet i Oslo.

Sommerville, I. (2000): *Software Engineering*, 6th ed. Edition Addison –Wesley.

Statskontoret (2000): *24-timmarsmyndighet: Förslag till kriterier för statlig elektronisk förvaltning i medborgarnas tjänst*, Rapport 2000:21, Stockholm

Videnskapsministeriet (2003): *E-service på borgernes præmisser? Statusrapport for Bedst på Nettet 2002*, IT- og Telestyrelsen, <http://www.bedstpaanettet.dk/status2002>

Ølnes, Svein (2001): *Oppsummering av kvalitetsevaluering 2001*, Vestlandsforskning, VF-notat 7/2001. Sogndal

Ølnes, Svein (2003): *Kvalitesevaluering av offentlege vevtenester*, Vestlandsforskning , VF-rapport 9/2003. Sogndal.

Ølnes, Svein og Terje Aaberge: *Enklare tilgang til offentlege informasjon - evaluering av LivsIT*, Vestlandsforskning, VF-rapport 7/2003. Sogndal.